日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: ２００４年　２月２４日

出願番号
Application Number: 特願２００４－０４７１７６

[ST. 10/C]: [ＪＰ２００４－０４７１７６]

出願人
Applicant(s): 株式会社日立製作所

２００４年　３月２４日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫

出証番号　出証特２００４－３０２４１８６

【書類名】　特許願
【整理番号】　K03011801A
【あて先】　特許庁長官殿
【国際特許分類】　G06F 12/00
【発明者】
　【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所システム開発研究所内
　【氏名】　出射　英臣
【発明者】
　【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所システム開発研究所内
　【氏名】　西川　記史
【発明者】
　【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所システム開発研究所内
　【氏名】　茂木　和彦
【特許出願人】
　【識別番号】　000005108
　【氏名又は名称】　株式会社　日立製作所
【代理人】
　【識別番号】　100075096
　【弁理士】
　【氏名又は名称】　作田　康夫
【選任した代理人】
　【識別番号】　100100310
　【弁理士】
　【氏名又は名称】　井上　学
【国等の委託研究の成果に係る記載事項】　国等の委託研究の成果に係る特許出願（平成１５年度　文部科学省　科学技術試験研究委託費〔１〕差分データ及びログ転送によるリモートサーバレス化技術の実現方式検討、２）クエリプラン利用先読み技術の実現方式検討、３）モニタ情報に基づくデータベース管理の自律化技術の実現方式検討）（産業活力再生特別措置法第３０条の適用を受けるもの）
【手数料の表示】
　【予納台帳番号】　013088
　【納付金額】　21,000円
【提出物件の目録】
　【物件名】　特許請求の範囲　1
　【物件名】　明細書　1
　【物件名】　図面　1
　【物件名】　要約書　1

【書類名】特許請求の範囲
【請求項１】
第一の計算機及び第一のストレージ装置を有する第一のサイト、
第二の計算機及び第二のストレージ装置を有する第二のサイト、及び
管理用の計算機、前記第一のサイト、前記第二のサイト及び前記管理用の計算機とを相互に接続するネットワークとを有し、
前記第一のストレージ装置は、前記管理用の計算機に入力された情報に基づいて該ストレージ装置が有する記憶領域に格納されるデータをグルーピングし、前記グルーピングされたグループ単位で該グループ内で更新されたデータを前記第二のストレージ装置へ転送し、
前記第一のサイトが停止した際に、
前記第二のサイトが前記グループ単位にデータを復旧することを特徴とする計算機システム。
【請求項２】
前記第一のストレージ装置は、
前記第二のサイトでデータを復旧する際に要求される復旧時間に基づいて優先順位をつけて前記グルーピングを行い、
前記第二のサイトは、
前記要求される復旧時間の優先順位が高い順に前記グルーピングされたグループに含まれるデータを復旧することを特徴とする請求項１記載の計算機システム。
【請求項３】
前記第一のストレージ装置から前記第二のストレージ装置への前記グループ単位のデータ転送は非同期リモートコピーで行われることを特徴とする請求項２記載の計算機システム。
【請求項４】
前記グルーピングされたデータは、データベースに使用されるデータであることを特徴とする請求項３記載の計算機システム。
【請求項５】
前記グルーピングされるデータには、前記データベースで使用されるログデータが含まれており、
前記ログデータが含まれるグループのデータは、同期リモートコピーで前記第一のストレージ装置から前記第二のストレージ装置へ転送されることを特徴とする請求項４記載の計算機システム。
【請求項６】
前記ログデータが含まれるグループには一番高い復旧時間の優先順位が割り当てられていることを特徴とする請求項５記載の計算機システム。
【請求項７】
前記第二のサイトは、前記第一のサイトの停止を前記ネットワークを介して検知することを特徴とする請求項６記載の計算機システム。
【請求項８】
前記第二のサイトは、
前記グループ単位でデータを復旧する際に、復旧前のデータが含まれているグループの使用を禁止し、グループ単位でのデータの復旧が完了するたびに復旧されたデータが含まれるグループの使用を許可することを特徴とする請求項７記載の計算機システム。
【請求項９】
計算機及び管理用計算機と接続されるストレージ装置であって、
他のストレージ装置とネットワークを介して接続され、
制御部及びメモリを有し、
前記制御部は、前記管理用計算機に入力された情報に基づいて該ストレージ装置が有する記憶領域に格納されるデータをグルーピングし、前記グルーピングの情報を前記メモリ

に格納し、
前記メモリに格納された情報に基づいて、前記グルーピングされたグループ単位で該グループ内で更新されたデータを前記第二のストレージ装置へ転送することを特徴とするストレージ装置。
【請求項１０】
前記グルーピングは、前記他のストレージ装置におけるデータの復旧時間に基づいた優先順位に基づいて行われることを特徴とする請求項９記載のストレージ装置。
【請求項１１】
第一の計算機及び第一のストレージ装置を有する第一のサイト、
第二の計算機及び第二のストレージ装置を有する第二のサイト、及び
管理用の計算機、前記第一のサイト、前記第二のサイト及び前記管理用の計算機とを相互に接続するネットワークとを有し、
前記第一のストレージ装置は、前記管理用の計算機に入力された前記第二のサイトでデータを復旧する際に要求される復旧時間に基づいた優先順位に基づいてストレージ装置が有する記憶領域に格納されるデータをグルーピングし、前記グルーピングされたグループ単位で該グループ内で更新されたデータを前記第二のストレージ装置へ転送し、
前記第一のサイトが停止した際に、
前記第二のサイトは、
前記要求される復旧時間の優先順位が高い順に前記グルーピングされたグループに含まれるデータを復旧することを特徴とする計算機システム。

【書類名】明細書
【発明の名称】計算機システム
【技術分野】
【０００１】
本発明は、計算機システム、特にデータベースを運用する計算機システム（以下「ＤＢシステム」）において、通常運用されているＤＢシステム（以下「現用系システム」）と、現用系システムのデータが複製されて使用される計算機システム（以下「待機系システム」）間におけるデータコピー方法及び待機系システムにおけるデータ回復の方法に関する。
【背景技術】
【０００２】
データベース（以下「ＤＢ」とも称する）が構築され、そのＤＢを運用しているＤＢシステムにおいて、ＤＢを構成するデータ（以下「ＤＢデータ」）のバックアップを取る方法として、現用系システムと同じ構成のＤＢシステムを待機系システムとして用意し、現用系システムのＤＢデータを待機系システムにコピーする技術がある。
【０００３】
上記方法を用いれば、現用系システムが何らかの理由により稼動を停止した際、待機系システムを稼動させることでシステム全体の可用性が向上する。また、現用系システムと待機系システムとを物理的に距離が離れた場所に各々設置することで、地震等の自然災害が現用系システムを設置した場所に発生した場合でも、待機系システムでＤＢデータを回復することが可能となる。以下、このような現用系、待機系システムを有するシステムをディザスタリカバリ（ＤＲ）システムと呼ぶ。
【０００４】
ＤＲシステムにおいて、現用系システムと待機系システムとの間のデータ転送が、ＤＢデータを格納するストレージ装置間で行われる場合がある。このようなストレージ装置間のデータ転送を以下リモートコピーと称する。
【０００５】
リモートコピーの手法として、同期リモートコピーと非同期リモートコピーの２種類がある。同期リモートコピーでは、現用系システムのストレージ装置と待機系システムのストレージ装置間でＤＢデータの更新の同期を取る（ＤＢデータの内容の同一性を保証する）ため、待機系システムに転送されたデータの信頼性は高い。一方、ＤＢデータの同期をとるため、現用系システムはＤＢデータを待機系システムへ転送している間処理の終了の報告を待つ必要があり、現用系システムの性能が低下する。
【０００６】
一方、非同期リモートコピーでは、現用系システムのストレージ装置と待機系システムのストレージ装置間でＤＢデータの同期を取らずにデータを転送するため、同期リモートコピーとは逆に、現用系システムの性能はそれほど低下しないが、待機系システムに転送されたデータの信頼性（あるいは最新性）は低くなる。
【０００７】
特許文献１では、現用系と待機系のストレージ装置間において、論理ボリュームグループ毎に優先度をつけ、その優先度に従って非同期リモートコピーを実行する方法を開示している。また、その一例としてＤＢのログが記憶される論理ボリュームの優先度を高くし、他のＤＢデータと比して、ログを優先的に待機系システムのストレージ装置へ転送することが開示されている。
【０００８】
【特許文献１】特開２００３－６０１６号公報
【発明の開示】
【発明が解決しようとする課題】
【０００９】
上述のＤＲシステムを構築する際、ＤＢデータの種類毎に、待機系システムにおけるＤ

Ｂデータの復旧に掛けられる時間が異なる場合がある。しかし従来技術ではその復旧時間の差が考慮されておらず、通常、最優先に復旧処理を行わなければならないＤＢデータの復旧時間内で全てのＤＢデータの復旧処理を行うようなＤＲシステムが構築される。しかしこれではＤＲシステムがオーバースペックになる可能性が高く、最適なシステム構成とはならない。

【００１０】

本発明の目的は、最優先に復旧すべきデータ、例えばＤＢデータの待機系システムにおける復旧時間を短く保ちながら、効果対価格比の良いシステムを構築することにある。

【課題を解決するための手段】

【００１１】

上記目的を達成するため、本発明では以下の構成を採用する。具体的には、正サイトから副サイトへデータをリモートコピーする際に、予め副サイトにおいて要求されるデータの復旧時間に応じて正サイトでデータを区別（グループ化）し、その区別に応じてデータを転送する順序を決定し、正サイトがリモートコピーを行う。その後、副サイトにおいては、正サイトの障害発生時に、上述したグループに基づいてデータの復旧を行う構成とする。

【００１２】

データのグループ化を行う際には、副サイトにおいてデータ復旧に要求される時間が短い順にグループ化を行い、正サイトはそのグループに基づいてデータを転送し、かつ副サイトにおいてデータ復旧にかけられる時間が少ないグループ順に復旧を行う。

【００１３】

さらに、データ復旧にかけられる時間の多少に関わらず、データベースにおけるログデータについては最先にデータを副サイトに転送してもよい。さらにこの場合、同期リモートコピーを用いてログデータを副サイトに転送しても良い。

【００１４】

又、副サイトにおいてデータを復旧する際に、一旦全てのデータを閉塞し、復旧されたグループ単位で閉塞を解除しても良い。

【発明の効果】

【００１５】

本発明によれば、待機系システムにおいて重要度の高いデータの復旧時間を短くしながらより最適なＤＲシステムを構築することが可能となる。

【発明を実施するための最良の形態】

【００１６】

以下、図１～図１４を用いて本発明の実施の形態を説明する。なお、これにより本発明が限定されるものではない。

図１は、本発明を適用した計算機システムの実施形態の例を示す図である。

計算機システムは、現用系システムである正サイト１００、待機系システムである副サイト１０２、正サイト１００と副サイト１０２とを接続するネットワーク１７２及び正サイト１００と副サイト１０２を管理するために使用されるシステム管理サーバ１５０を有する。

【００１７】

正サイト１００及び副サイト１０２は、データベースマネージメントシステム（以下、ＤＢＭＳ）が動作するＤＢサーバ１１０、ＤＢデータを記憶するストレージ装置１３０並びにＤＢサーバ１１０及びストレージ装置１３０を相互に接続するネットワーク１７０を有する。以下、正サイト１００と副サイト１０２の機器構成は同等として説明するが、全く同じでなくても良い。少なくとも双方のサイトでＤＢサーバとストレージ装置があれば良い。

【００１８】

ＤＢサーバ１１０は一般的な計算機であり、ユーザが使用する計算機もしくは当該計算機が接続されるネットワークとのインタフェースであるＩ／Ｆ（Ａ）１１２、制御装置（

制御プロセッサ）１１４、入出力装置１１６、メモリ１１８、ネットワーク１７０とのインタフェースであるＩ／Ｆ（Ｂ）１２４を有する。また、メモリ１１８には、運用管理プログラム１２０及びＤＢＭＳを実行するプログラム（以下「ＤＢＭＳ」と略する）１２２が格納される。

【００１９】

ストレージ装置１３０は、ネットワーク１７０とのインタフェースであるＩ／Ｆ１３２、制御装置（制御プロセッサ）１３４、メモリ１３６及びディスク１４４を有する。また、メモリ１３６には、制御プログラム１３８、リモートコピー（以下「ＲＣ」）を実行するプログラム１４０及び制御情報１４２が格納される。尚、ディスク１４４は記憶媒体であり、ハードディスクドライブでも、光ディスクでも良い。また、ストレージ装置１３０は、複数のディスク１４４をＲＡＩＤ構成にしていても良い。

【００２０】

計算機システムの管理者は、正サイト１００内のネットワーク１７０に接続されたシステム管理サーバ１５０により、計算機システムの設定・管理を行う。

【００２１】

システム管理サーバ１５０は一般的な計算機であり、制御装置（制御プロセッサ）１５２、入出力装置１５４、メモリ１５８及び正サイトのネットワーク１７０とのインタフェースであるＩ／Ｆ１６０を有する。なお、システム管理サーバ１５０と正サイトのＤＢサーバ１１０とを一体とした構成も考えられる。この場合、管理プログラム１５８がＤＢサーバ１１０で実行される。

【００２２】

又、上述した各種プログラムは、光ディスクやフロッピー（登録商標）ディスク等の記憶媒体を用いて各サイトにインストールされたり、ネットワークを介して各サイトの各装置にインストールされる。尚、以下でプログラムが主語になる場合は、実際はそのプログラムを実行する制御装置が処理を実行している。

【００２３】

図２の（a）は、ストレージ装置１３０が持つ制御情報１４２の構成例を示した図である。

制御情報１４２は、ストレージ装置１３０が有する論理的な記憶領域（論理ユニット：ＬＵ）の設定に関する情報であるＬＵ情報２００及びデータの優先度毎にグルーピングされるグループの情報であるグループ情報２０２を有する。尚、ＬＵは、一つ又はそれ以上のディスク１４４から構成される。

【００２４】

図２の（ｂ）は、ＬＵ情報２００の構成例を示す図である。ＬＵ情報２００は、各ＬＵ毎に対応するエントリを有する。各エントリは、エントリに対応するＬＵを識別する番号であるＬＵＮを登録するフィールド２１０、対応するＬＵに割当てられたブロックの数を示すブロック数を登録するフィールド２１２及び対応するＬＵが割当てられるＤＢサーバ１１０を識別するためのサーバＩＤを登録するフィールド２１４を有する。なお、ブロックとは記憶領域を扱う際の単位を示し、通常512KBである。

【００２５】

図２の（ｃ）は、グループ情報２０２の構成例を示す図である。グループ情報２０２は、優先度のグループ毎に対応するエントリを有する。各エントリは、エントリに対応するグループを識別するグループＩＤが登録されるフィールド２２０、対応するグループ内のデータの優先度を示すデータ優先度が登録されるフィールド２２２及び対応するグループに属するＬＵを示すＬＵＮを登録するフィールド２２４を有する。

【００２６】

図３は、運用管理プログラム１２０が保持するローデバイス情報３００の例を示した図である。このローデバイス情報３００は、ＤＢサーバ１１０のメモリ１１８に格納されている。

【００２７】

ローデバイス情報３００は、ＤＢサーバ１００に登録されるローデバイス毎に対応するエントリを有する。ここでローデバイスとは、ＤＢサーバ１００で実行されるオペレーティングシステム、特にファイルシステムにおいて認識される仮想的な外部装置を指す。ファイルシステムはこのローデバイスを１つのファイルとして表現して操作する。これをローデバイスファイルと言う。

各エントリは、該エントリに対応するローデバイスファイルのファイル名を登録するフィールド３０２、ローデバイスファイルで表現されるローデバイスと関連付けられるＬＵを有するストレージ装置１３０の識別子を登録するフィールド３０４及びローデバイスに対応するＬＵのＬＵＮを登録するフィールド３０６とを有する。

【００２８】

図４の（ａ）は、ＤＢＭＳ１２２の構成例を示した図である。

ＤＢＭＳ１２２は、ＤＢＭＳ実行プログラム４００、ＤＢデータの読み出しや書き込みの際にＤＢデータを一時保持するＤＢバッファ４０２、ＤＢデータの更新に伴って追加されるログを一時保持するログバッファ４０４、ＤＢＭＳ１２２のシステム情報やログ、ＤＢデータ等を記憶させるデータ領域の設定情報であるデータ領域設定情報４０６及び表や索引といったＤＢＭＳ１２２のスキーマ情報であるＤＢスキーマ情報４０８を有する。

【００２９】

図４の（ｂ）は、データ領域設定情報４０６の構成例を示す図である。データ領域設定情報４０６は、データ領域毎に対応するエントリを有する。各エントリは、該エントリに対応するデータ領域を識別するためのデータ領域ＩＤを登録するフィールド４１０、対応するデータ領域の名前を示すデータ領域名を登録するフィールド４１２、対応するデータ領域が作成されるローデバイスファイルのファイル名が登録されるフィールド４１４、対応するデータ領域に格納されるデータの種類を示すデータ種別が登録されるフィールド４１６、対応するデータ領域の大きさを示す領域サイズが登録されるフィールド４１８及び対応するデータ領域に格納されるデータの優先度を示すデータ優先度を登録するフィールド４２０を有する。

【００３０】

図４の（ｃ）は、ＤＢスキーマ情報４０８の構成例を示す図である。ＤＢスキーマ情報４０８は、ＤＢスキーマ毎に対応するエントリを有する。各エントリは、エントリに対応するＤＢスキーマを識別するためのスキーマＩＤを登録するフィールド４３０、対応するＤＢスキーマの名前を示すスキーマ名を登録するフィールド４３２、対応するＤＢスキーマの種類を示すスキーマ種別を登録するフィールド４３４、対応するＤＢスキーマが作られるデータ領域を識別するデータ領域ＩＤを登録するフィールド４３６及び対応するＤＢスキーマに割り当てている記憶領域のデータサイズを示すスキーマサイズを登録するフィールド４３８を有する。

【００３１】

なお、スキーマ種別が“ＴＡＢＬＥ”の場合はＤＢの表データ、“ＩＮＤＥＸ”の場合はＤＢの索引データがそのスキーマ内に格納される。

【００３２】

図５は、ＤＢＭＳ１２２によってストレージ装置１３０に作成されるデータ領域の構成例を示した図である。上述したように、データ領域はデバイスファイル名を介してストレージ装置１３０が有する各ＬＵと対応付けられる。

データ領域５００は、データ領域の管理情報を格納するデータ格納領域管理領域５０２及び実際のＤＢスキーマのデータを格納するデータ格納領域５０４を有する。データ領域管理領域５０２は、データ領域を識別するためのデータ領域ＩＤ５１０を格納する領域を有する。

【００３３】

図５（ａ）は、データ領域５００にログデータを格納した例を示した図である。ログデータはデータ格納領域５０４に格納される。ログデータが格納される領域５２０には、複数のログエントリが格納される。一つのログエントリの領域は、対応するログエントリを

識別するために付与されるログシーケンス番号（以下、ＬＳＮ）を格納する領域５３０、対応するログの種類を識別するログ種別を格納する領域５３２、追加更新されたデータが記憶されているデータ領域を示すデータ領域ＩＤが格納される領域５３４、ログエントリに対応するデータのアドレスが格納される領域５３６及び追加更新されたデータ内容が格納される領域５３８を有する。

【００３４】

このログデータの１つ１つのエントリは、ＤＢへデータが追加更新された時や、ＤＢがコミットされた際にＤＢＭＳ１２２によって自動的に作成される。作成された後、このログエントリは一旦ログバッファ４０４に保持される。そして、ログバッファ４０４の記憶容量が不足した時や、ある決められた時間が経過した時、ＤＢＭＳ１２２はログバッファ４０４に保持されているログデータをストレージ装置１３０のデータ領域５００に書き込む。

【００３５】

図５（ｂ）は、データ領域にＤＢデータを格納した場合を示した図である。ＤＢデータはデータ格納領域５０４に格納される。ＤＢＭＳ１２２がＤＢデータを追加又は更新した際、追加更新のＤＢデータは一旦ＤＢバッファ４０２に保持される。そして、ＤＢバッファの記憶容量が不足した時や、ある決められた時間が経過した時、ＤＢＭＳ１２２はＤＢバッファ４０２に保持されているＤＢデータをストレージ装置１３０のデータ領域に書き込む。

【００３６】

また、データ領域管理領域５０２は、そのデータ領域内で最も最近更新されたデータのＬＳＮを格納する領域５１２を有する。例えば、あるデータ領域のデータが更新された時に作られたログデータのＬＳＮが１０の場合、この１０という値が領域５１２に格納される。これにより、ＤＢサーバ１１０が、ＬＳＮ１０までのログデータが当データ領域に反映されていることを判断することが可能となる。

【００３７】

図６、図７は、計算機システムにおける正サイト１００から副サイト１０２へのデータ転送の概略を示した図である。

ストレージ装置１３０は、ＲＣプログラム１４０を実行して、正サイト１００から副サイト１０２へデータをコピーする。このリモートコピーは、グループ単位で優先度順に行われる。また、優先度（「データ優先度」とも言う）が０であるグループのデータ（本実施形態ではログデータ）は同期リモートコピー、データ優先度が０以外のグループのデータ（本実施の形態ではＤＢデータ）は非同期リモートコピーで転送される。

【００３８】

図６において、正サイト１００のグループ０（６００）はデータ優先度が０であるため、このグループ０（６００）に属するＬＵ内のデータが更新される度に、正サイト１００のストレージ装置１３０は、同期リモートコピー（矢印６１０）によって副サイト１０２のストレージ装置１３０にデータを転送する。

【００３９】

一方、図６においてグループ１（６０２）、グループ２（６０４）、グループ３（６０６）のデータ優先度は、それぞれ１、５、３である。従って、これらのグループに属するＬＵ内のデータを副サイト１０２にコピーする際、正サイト１００のストレージ装置１３０は、グループ１（６０２）に属するＬＵのデータ、グループ３（６０６）に属するＬＵのデータ、グループ２（６０４）に属するＬＵのデータの順番で非同期リモートコピー（矢印６１２、６１４、６１６）で副サイト１０２のストレージ装置１３０へデータ転送を行う。

【００４０】

図７は、データ領域の観点から見た正副サイト間のデータ転送の概略を示した図である。以下、正サイト及び副サイトのストレージ装置１３０に、グループ１、２及び３に対応するデータ領域７００、７０２及び７０４が構成されているとする。尚、データ領域７０

０等は上述したデータ領域５００の具体例であるとする。又、グループは同じ優先度に属するＬＵで構成される。

ＤＢＭＳ１２２によって、データ領域７００にデータ「Ａ」、データ領域７０２にデータ「Ｂ」、データ領域７０４にデータ「Ｃ」、データ領域７００にデータ「Ｄ」、データ領域７０２にデータ「Ｅ」及びデータ領域７０４にデータ「Ｆ」という順番でデータ更新が行われた場合、図７のログデータ７１０（データ領域５００の一種）に示されるログが作られる。

【００４１】

ログデータ７１０を格納するデータ領域は、データ優先度が０のグループに属しているＬＵに作られているため、データの書き込みがある度にそれらのデータは同期リモートコピー（矢印６１０）で副サイト１０２へ転送される。

【００４２】

データ領域７００、データ領域７０２及びデータ領域７０４は、それぞれ前述のグループ１（６０２）、グループ２（６０４）及びグループ３（６０６）に属しているＬＵに作られているため、データ優先度に従い、データ領域７００のデータ「Ａ」「Ｄ」、データ領域７０４のデータ「Ｂ」「Ｅ」、データ領域７０２のデータ「Ｃ」「Ｆ」という順番で、非同期リモートコピーで正サイト１００から副サイト１０２にデータが転送される。また、データが正サイト１００から副サイト１０２へコピーされた際は、そのデータに対応するログのＬＳＮも、副サイト１０２のデータ領域管理領域５０２のＬＳＮ５１２に転送される。

【００４３】

図８～図１４は、計算機システムにおける、記憶領域の設定処理から災害発生によるシステム回復処理までの一連の処理手順の例を示したフロー図である。

簡略に処理手順を説明すると、まず、計算機システムの管理者等は計算機システムが使用する記憶領域の設定を行う。この際、管理者等は副サイトでのデータの復旧時間に基づく優先順位に基づいて、記憶領域の優先度を決め、その情報を正ストレージ装置や副サイトへ登録する。その後、正副サイトは、設定された優先度に基づいてデータのリモートコピーを行う。

【００４４】

正サイトで障害が発生した場合、管理者は副サイトでシステムの復旧を図る。この際、副サイトの計算機システムは、あらかじめ設定された優先度の順番にデータの復旧処理を行う。この際、副サイトのストレージ装置は、一旦すべての記憶領域の使用を禁止し、その後、復旧が完了した記憶領域から順に使用禁止を解除していく。以下、各段階での処理手順を詳細に説明する。

【００４５】

図８は、記憶領域の割当て処理の手順例を示したフロー図である。

計算機システムの管理者は、計算機システムを構築するにあたり、まずシステム管理サーバ１５０の管理プログラム１５８を起動し、正サイト１００のストレージ装置１３０内のＬＵを設定する。ここで設定された内容は、ネットワークを介してストレージ装置１３０が有するＬＵ情報２００に格納される。具体的には、例えばＬＵ情報２００のエントリ２１６は、管理者がＬＵＮが０のＬＵに20971520ブロックを割り当て、かつサーバＩＤが０のＤＢサーバ１１０にＬＵＮ０のＬＵを割り当てたことを示している（ステップ８０２）。

【００４６】

続いて、システム管理者は、正サイト１００のＤＢサーバ１１０上のローデバイスを設定する。ここで設定された内容は、ネットワークを介してＤＢサーバ１１０が有するローデバイス情報３００に格納される。具体的には、例えばローデバイス情報３００のエントリ３０８は、管理者が“/dev/rdsk/c0t0d1s6”というファイル名のローデバイスに、ストレージ装置ＩＤ０のストレージ装置１３０が有するＬＵＮ０のＬＵを対応付けた（以下「マッピング」）ことを示している。なお、ローデバイスにマッピング可能なＬＵは、ステ

ップ８０２で設定したＬＵである（ステップ８０４）。

【００４７】

続いて管理者は、副サイト１０２のストレージ装置１３０内のＬＵを設定する。尚、設定の内容は、サーバＩＤ以外ステップ８０２で設定した内容と同じにする必要がある。これは、リモートコピー時の正副サイト間の整合性をとるためである（ステップ８０６）。

【００４８】

続いて管理者は、副サイト１０２のＤＢサーバ１１０のローデバイスを設定する。設定の内容は、ストレージ装置ＩＤ以外ステップ８０４で設定した内容と同じにする必要がある。これも、正副サイト間の整合性を取るためである（ステップ８０６）。

【００４９】

図９は、計算機システムにおけるＤＢシステムの構築処理の手順例を示したフロー図である。本処理は、記憶領域の割り当て処理の後に行われる。

ＤＢの構築を行うＤＢ設計者等は、正サイト１００のＤＢＭＳ１２２を起動し、データ領域の名前、データ領域を作成するローデバイスのローデバイスファイル名、データ領域に格納するデータの種別、データ領域のサイズ、格納するデータの優先度を指定したクエリを入出力装置１１６から入力してデータ領域を作成する。

【００５０】

データの優先度は具体的には以下のように決定する。まず、ログデータは副サイト１０２におけるデータ復旧に欠かせないため、ログデータを格納するデータ領域のデータ優先度は最も高く設定する。それ以外のＤＢデータは、データの重要度やＤＢ処理の内容によって、障害発生時、副サイト１０２において早く復旧してほしい順にデータ優先度を決定する。例えば、常々更新処理が行われるオンライン系のＤＢデータは一刻も早く復旧して業務を再開する必要があるため、データ優先度を高く設定する。夜間などに一括して処理するバッチ系のＤＢデータは決められた時間内で処理されるため、優先度は中ほどに設定する。アーカイブ系のＤＢは検索／更新処理が頻繁に発生しないため、優先度は低く設定する。

【００５１】

データ領域を作成した時点で、ＤＢＭＳ１２２は、そのデータ領域に識別子（ＩＤ）を自動的に割り振る。ここで設定された内容は、ＤＢＭＳ１２２内のデータ領域設定情報４０６に格納される。具体的には、例えばデータ領域設定情報４０６のエントリ４２２は、データ領域ＩＤに０が割り振られたデータ領域の領域名は“ＬＯＧ”であり、“/dev/rdsk/c0t0d1s6”というローデバイスファイル上に作られていることを示す。又、このデータ領域に格納されるデータはＤＢＭＳ１２２のログデータで、領域の大きさは４ＧＢ、データの優先度は０であることを示している（ステップ９０２）。

【００５２】

ステップ９０２でデータ領域が作られた後、ＤＢサーバ１１０はデータ領域設定完了通知と一緒にローデバイス情報３００及びデータ領域設定情報４０６をストレージ装置１３０に送信する。ストレージ装置１３０は、ＤＢサーバ１１０からデータ領域設定完了通知、ローデバイス情報３００及びデータ領域設定情報４０６を受信した後に、制御プログラム１３８のグループ作成処理１０００を実行する。グループ作成処理１０００については後述する。（ステップ９０４）。

【００５３】

続いて、ＤＢ作成者は、ステップ９０２で作成したデータ領域上に、ＤＢスキーマの名前、種別、ＤＢスキーマを作成するデータ領域のＩＤ、ＤＢスキーマのサイズをＤＢＭＳ１２２に指定してＤＢスキーマの作成を指示する。ＤＢスキーマを作成した時点で、ＤＢＭＳ１２２はそのＤＢスキーマにスキーマＩＤを自動的に割り振る。ここで設定された内容は、ＤＢＭＳ１２２内のＤＢスキーマ情報４０８に格納される。

【００５４】

具体的には、例えばＤＢスキーマ情報４０８のエントリ４４０は、ＤＢスキーマＩＤ０のスキーマ名は“ＬＯＧ１”であり、データ領域ＩＤが０のデータ領域に作られているこ

とを示している。又、このＤＢスキーマはＤＢＭＳのログデータを格納するスキーマであり、スキーマの大きさは２ＧＢであることを示している。

尚、一つのＤＢＭＳ１２２によってログが複数発生する場合、例えばＤＢＭＳ１２２が２つ以上のログを切り替えて使用するといった場合は、ログ用のデータ領域を複数作成することで対応する（ステップ９０６）。

【００５５】

続いて、ＤＢ作成者は、副サイト１０２のＤＢＭＳ１２２を起動し、ステップ９０２で作成したデータ領域と同じデータ領域を副サイト１０２にも作成する（ステップ９０８）。

【００５６】

ステップ９０８で副サイト１０２にデータ領域が作られた後、副サイト１０２のストレージ装置１３０は、制御プログラム１３８のグループ作成処理１０００を実行する。グループ作成処理１０００については後述する（ステップ９１０）。

【００５７】

続いて、ＤＢ作成者は、ステップ９０８で作成したデータ領域に、ステップ９０６で作成したＤＢスキーマと同じＤＢスキーマを作成する。なお、ステップ９０８から本ステップにおいて正サイト１００と同じ処理を繰り返す代わりに、正サイト１００のデータ領域設定情報４０６や、ＤＢスキーマ情報４０８を副サイト１０２へコピーして、正サイト１００と同じＤＢを副サイト１０２上に構築しても良い（ステップ９１２）。

その後、正サイト１００のＤＢＭＳ１２２の運用を開始する（ステップ９１４）。

【００５８】

図１０は、制御プログラム１３８によるグループ作成処理１０００の処理手順の例を示したフロー図である。なお、上述したように本処理はデータ領域が作成されたときに実行される。

ストレージ装置１３０は、制御プログラム１３８を実行して、作られたデータ領域のデータ優先度とそのデータ領域が属するＬＵのＬＵＮについての情報をＤＢサーバ１１０のローデバイス情報３００及びデータ領域設定情報４０６から取得し、そのＬＵＮに対応するＬＵのデータ優先度を決定する（ステップ１００２）。

【００５９】

続いて、ストレージ装置１３０は、ステップ１００２で取得したデータ優先度と同じデータ優先度のグループがあるか、グループ情報２０２を検索する（ステップ１００４）。

【００６０】

ステップ１００４で同じデータ優先度のグループがあった場合、ストレージ装置１３０は、同じデータ優先度のグループにステップ１００２で取得したＬＵＮを追加し、グループ情報２０２を更新する（ステップ１００６）。

【００６１】

一方、ステップ１００５で同じデータ優先度のグループがなかった場合、ストレージ装置１３０は、グループ情報２０２に新たなエントリを追加することで新しいグループを作成し、ステップ１００２で取得したＬＵＮをそのエントリのフィールド２４４に追加する（ステップ１００８）。

【００６２】

上述のステップによって、グループ情報２０２が更新される。具体的には、例えばグループ情報２０２のエントリ２２６は、グループＩＤが０のグループ内のデータ優先度は０であり、ＬＵＮ０のＬＵが当グループに属していることを示している。

【００６３】

図１１は、ＤＢＭＳ１２２によるデータ更新処理の手順例を示したフロー図である。

正サイト１００のＤＢＭＳ１２２は、ユーザからのＤＢ処理（データ更新）要求により、更新するデータをストレージ装置１３０からＤＢバッファ４０４に読み出してデータを更新する（ステップ１１０２）。

【００６４】

続いてＤＢＭＳ１２２は、当データ更新を識別するために付与したＬＳＮ、ログの種別、更新されたデータのデータ領域ＩＤ、データのアドレス、更新データを一つのログとしてログバッファ４０２に書き込む（ステップ１１０４）。

【００６５】

続いてＤＢＭＳ１２２は、ログバッファ４０２の記憶領域に空きが無くなったかどうか判定する（ステップ１１０６）。ログバッファ４０２の記憶領域に空きが無い場合、ＤＢＭＳ１２２は、ログバッファ４０２内のログデータを、データ領域５００のうち、ログ格納用に作成したデータ領域に書き込む（ステップ１１０８）。

【００６６】

この際、ログ格納用のデータ領域のデータ優先度は０であるので、ストレージ装置１３０は、ＲＣプログラム１４０を実行して、ログデータを副サイト１０２へ同期リモートコピーする。尚、リモートコピー処理１２００については後述する（ステップ１１１０）。

【００６７】

続いて、ＤＢＭＳ１２２は、ＤＢバッファ４０４の記憶領域に空きがあるかどうか判定する（ステップ１１１２）。ＤＢバッファ４０４に空きが無くなった場合、ＤＢＭＳ１２２は、ＤＢバッファ４０４内の更新されたデータを、対応するデータ領域の所定のアドレスで示される場所に書き込む（ステップ１１１４）。又、ＤＢＭＳ１２２は、データ領域管理領域５０２のＬＳＮ５１２に、書き込んだ更新データに対応するログのＬＳＮを格納する（ステップ１１１６）。

【００６８】

なお、ログデータ及びＤＢデータのデータ領域への書き込み契機をログバッファ４０２及びＤＢバッファ４０４の記憶領域に空きが無くなった時で説明したが、各バッファの空きの有無ではなく、ある一定期間毎といった別の契機でＤＢＭＳ１２２がデータ領域への書き込みを行ってもかまわない。

【００６９】

図１２は、ＲＣプログラム１４０によるリモートコピー処理１２００の手順の例を示したフロー図である。

正サイト１００のＲＣプログラム１４０は、データ優先度が０のグループに属するＬＵのデータが更新された場合（ステップ１２０２）、更新されたデータを副サイト１０２に送信し（ステップ１２０４）、副サイト１０２からの応答を待つ（ステップ１２０６）。正サイト１００のＲＣプログラム１４０は、副サイト１０２からの応答を受信して次の処理に移る。

【００７０】

一方、正サイト１００でのデータ更新がデータ優先度０以外の場合又はデータ更新が発生しない場合、正サイト１００のＲＣプログラム１４０は決められたある一定時間の経過を待つ（ステップ１２０８）。一定時間経過後、正サイト１００のＲＣプログラム１４０は、データ優先度が０以外のグループに含まれるＬＵに対して行われたデータ更新を検索し、その更新されたデータをグループのデータ優先度順で副サイト１０２に送信する（ステップ１２１０）。

【００７１】

図１３は、副サイト１０２の運用管理プログラム１２０による災害監視処理の手順例を示したフロー図である。

副サイト１０２の運用管理プログラム１２０は、正サイト１００の運用管理プログラム１２０に対し、定期的にアライブ通信を行う。アライブ通信とは、お互いがお互いを監視しあうために遣り取りされる通信で、一般的にはハートビート信号等が使用される（ステップ１３０２）。

【００７２】

ステップ１３０２のアライブ通信に対する応答が正サイト１００から帰ってきた場合、副サイト１０２は、正サイト１００が正常に動作していると判定し、一定期間をおいてステップ１３０２からの処理を繰り返す。

【００７３】
一方、正サイト１００から応答がなかった場合、副サイト１０２は、正サイト１００に障害、もしくは災害が発生したと判定する。そして、副サイト１０２の運用管理プログラム１２０は、副サイト１０２のＤＢＭＳ１２２を起動して運用を開始するとともに（ステップ１３０６）、ＤＢＭＳ１２２は、ＤＢ回復処理１４００を実行する。なお、このときＤＢＭＳ１２２は、ＤＢデータを記憶する各データ領域を閉塞状態にしておき、データの回復が終わったデータ領域から順次オープンしていく。ここで「閉塞状態」とは、データ領域のデータにＤＢサーバ１１０のアプリケーションプログラム等がアクセスができない状態のことを指す。又、「オープン」とは、閉塞状態が解除され、データ領域のデータがアプリケーションプログラム等によって使用可能な状態にすることを指す（ステップ１３０８）。

【００７４】
図１４は、副サイト１０２のＤＢＭＳ１２２によるＤＢ回復処理１４００の手順例を示したフロー図である。
副サイト１０２のＤＢＭＳ１１２は、データ領域設定情報４０６を参照して、データ優先度が高いデータ領域からデータの回復を行い（ステップ１４０２）、データの回復が終わればそのデータ領域をオープンして使用可能状態にする（ステップ１４０４）
回復の方法の手順は次の通りである。まずＤＢＭＳ１２２は、データ領域のデータ領域管理領域５０２内にある領域５１２のＬＳＮを取得し、そのデータ領域で一番最後に更新されたデータのログを特定する。そして、ＤＢＭＳ１２２は、同期リモートコピーでコピーされたログデータから、ＬＳＮで特定したログより後のログの中でそのデータ領域のデータを更新したことを示すログがあれば、そのログのエントリに格納された更新データを用いてＤＢデータの更新を行う。

【００７５】
例えば、図７でデータ「Ａ」「Ｄ」「Ｃ」が副サイト１０２の各データ領域にコピーされている状態からデータ回復を行う場合、データ優先度が最も高いデータ領域７００のＬＳＮは「１３」であり、ＬＳＮが「１３」以降のログデータ７１０の中でデータ領域７００のデータ更新を示しているログはないため、ＤＢＭＳ１２２は、データ領域７００をそのままオープンして使用可能状態にする。

【００７６】
次にデータ優先度が高いデータ領域７０４のＬＳＮは「１２」であり、ＬＳＮが「１２」以降のログデータ７１０の中でＬＳＮが「１５」のログがデータ領域７０４のデータ更新を示している。従って、ＤＢＭＳ１２２は、ＬＳＮが「１５」のログが持つ更新データ「Ｆ」をデータ領域７０４に書き込み、領域５１２のＬＳＮを「１５」に更新した後、データ領域７０４をオープンして使用可能状態にする。

【００７７】
最後に最もデータ優先度が低いデータ領域７０２のＬＳＮ５１２は「９」であり、ＬＳＮが「９」以降のログデータ７１０の中でＬＳＮが「１１」と「１４」のログがデータ領域７０２のデータ更新を示している。従って、ＤＢＭＳ１２２は、ＬＳＮが「１１」のログが持つ更新データ「Ｂ」と、ＬＳＮ「１４」のログが持つ更新データ「Ｅ」をデータ領域７０２に書き込み、ＬＳＮ５１２を「１４」に更新した後、データ領域７０２をオープンして使用可能状態にする。

【００７８】
ここで、従来技術と本発明とのデータ回復の処理の違いを図７を用いて説明する。
図７において、更新のあったデータ「Ａ」～「Ｆ」が正サイトから副サイトにリモートコピーされる際、従来技術ではデータの更新順で副サイトにデータが転送されるが、本発明では各データが書き込まれるデータ領域の優先度順でコピーされる。つまり、データ「Ａ」～「Ｆ」の順番ではなく、優先度が最も高いグループ１のデータ「Ａ」「Ｄ」、次に優先度が高いグループ３のデータ「Ｃ」「Ｆ」、最後に優先度が最も低いグループ２のデータ「Ｂ」「Ｅ」の順番で副サイトにコピーされる。

【００７９】
ここで、４つ目のデータをコピーしている時に、正サイトに障害が発生し、副サイトでＤＢの復旧を図ると仮定する。従来技術ではデータ「Ａ」、「Ｂ」及び「Ｃ」が副サイトにコピーされた状態から復旧を図ることになり、ログデータ７１０からデータ「Ｄ」、「Ｅ」及び「Ｆ」を回復する。つまり、どのデータ領域にも回復が必要なデータが存在する可能性があり、ＤＢデータの復旧時間が延びてしまう。
【００８０】
一方、本発明ではデータ「Ａ」、「Ｄ」及び「Ｃ」が副サイトにコピーされた状態から、優先度の高いグループに属するデータ領域順に復旧を図るため、グループ１に属するデータ領域、グループ３に属するデータ領域、グループ２に属するデータ領域の順番でデータの回復を行う。
【００８１】
ここで本発明においては、グループ１に属するデータ領域７００のデータ「Ａ」及び「Ｄ」は既にコピーされているため、副サイト側のＤＢＭＳはログデータ７１０からＤＢデータを回復する必要はなく、グループ１に属するデータ領域をそのままオープンして使用可能状態にすることができる。又、グループ３に属するデータ領域７０４のデータ「Ｃ」は既にコピーされているため、副サイト側のＤＢＭＳはログデータ７１０からデータ「Ｆ」のみを回復した後、グループ３に属するデータ領域をオープンして使用可能状態にする。一方、グループ２に属するデータ領域７０２のデータ「Ｂ」「Ｅ」はまだコピーされていないため、副サイト側のＤＢＭＳはログデータ７１０からデータ「Ｂ」及び「Ｅ」を回復した後、グループ２に属するデータ領域をオープンして使用可能状態にする。以上の手順により、本発明では、優先度の高いデータ領域の復旧時間を短くすることが可能となる。
【００８２】
上述したように、本発明では、優先度を高く設定したデータ領域のデータを優先的に福サイトにコピーし、障害等により福サイト上でＤＢの回復を図る際は高優先度のデータ領域のデータから復旧を図ることで、優先度の高いデータ領域のデータの復旧時間を短くすることができる。
【００８３】
つまり、本発明においては、ある時点において正サイトに障害が発生し、副サイト上でＤＢ復旧が行われる際、復旧優先度の高いデータは復旧優先度の低いデータに比べ、副サイト上にコピーされている可能性が高くなるため、この場合ログからのデータ復旧を行う必要がなく、したがって復旧時間が短くなる。
【図面の簡単な説明】
【００８４】
【図１】計算機システムのシステム構成例を示す図である。
【図２】制御情報１４０の構成例を示した図である。
【図３】ローデバイス情報３００の例を示した図である。
【図４】ＤＢＭＳ１２２の構成例を示した図である。
【図５】データ領域の構成例を示した図である。
【図６】正サイト１００から副サイト１０２へのリモートコピーの概略を示した図である。
【図７】リモートコピーの概略を示した図である。
【図８】記憶領域の割当て処理の手順例を示したフロー図である。
【図９】ＤＢの構築処理の手順例を示したフロー図である。
【図１０】グループ作成処理の手順例を示したフロー図である。
【図１１】データ更新処理の手順例を示したフロー図である。
【図１２】リモートコピー処理の手順例を示したフロー図である。
【図１３】副サイト１０２の運用管理プログラム１２０による災害監視処理の手順例を示したフロー図である。

【図１４】副サイト１０２のＤＢＭＳ１２２によるＤＢ回復処理の手順例を示したフロー図である。

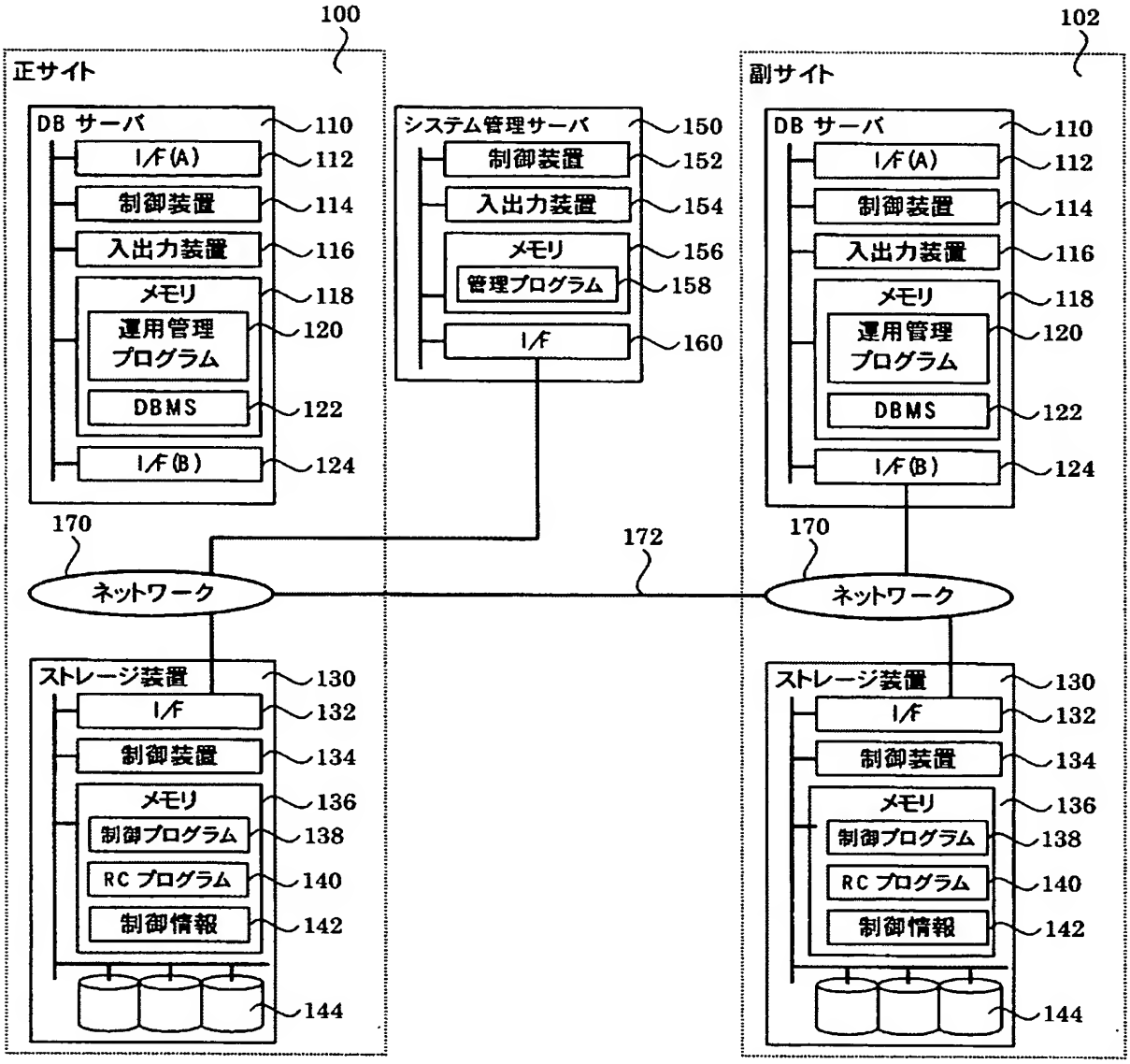
【符号の説明】

【００８５】

１００…正サイト、１０２…副サイト、１１０…ＤＢサーバ、１１２…Ｉ／Ｆ（Ａ）、１１４…制御装置、１１６…入出力装置、１１８…メモリ、１２０…運用管理プログラム、１２２…ＤＢＭＳ、１３０…ストレージ装置、１３４…制御装置、１３６…メモリ、１４４…ディスク、１５０…システム管理サーバ、１７０…ネットワーク。

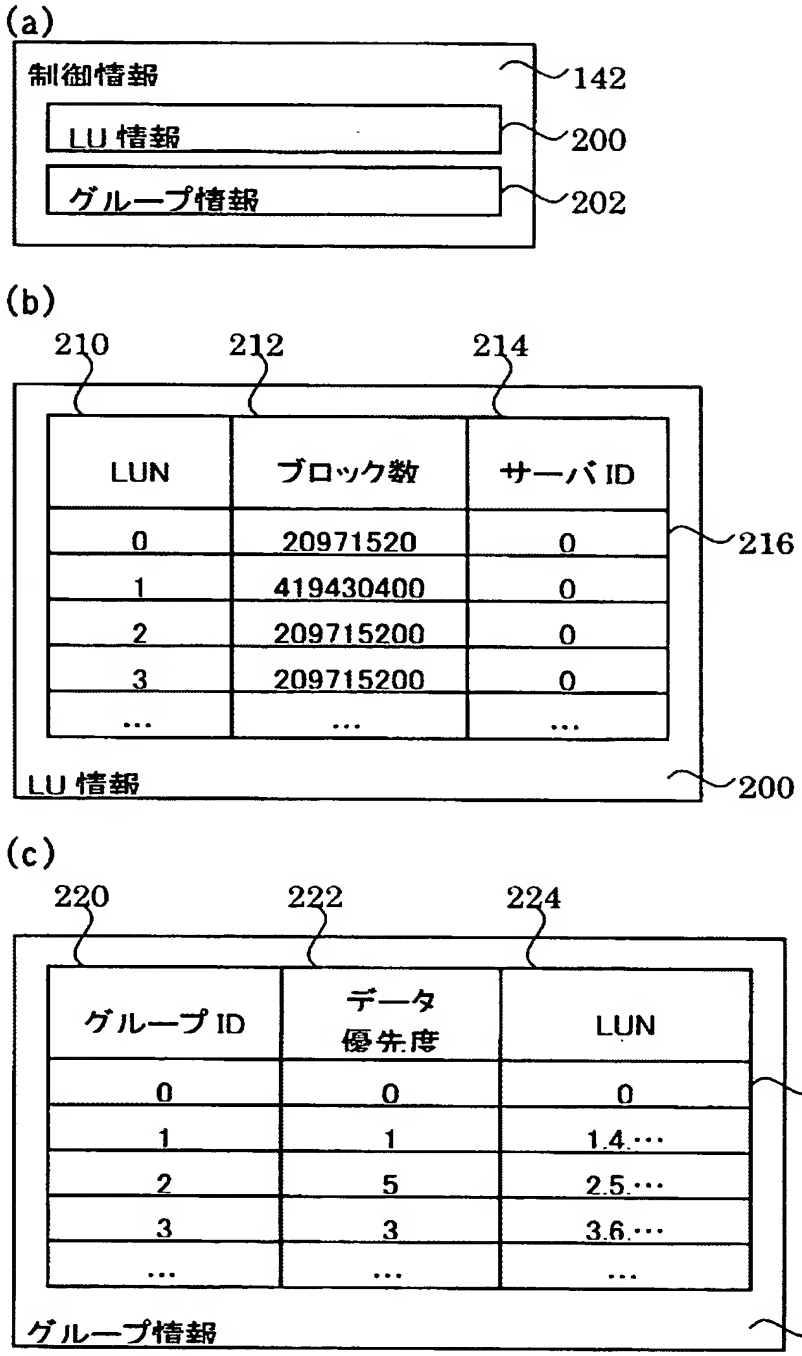
【書類名】図面
【図１】

図1


100
正サイト
DB サーバ
110
I/F(A)
112
制御装置
114
入出力装置
116
メモリ
118
運用管理プログラム
120
DBMS
122
I/F(B)
124
システム管理サーバ
150
制御装置
152
入出力装置
154
メモリ
156
管理プログラム
158
I/F
160
102
副サイト
DB サーバ
110
I/F(A)
112
制御装置
114
入出力装置
116
メモリ
118
運用管理プログラム
120
DBMS
122
I/F(B)
124
170
ネットワーク
172
170
ネットワーク
ストレージ装置
130
I/F
132
制御装置
134
メモリ
136
制御プログラム
138
RC プログラム
140
制御情報
142
144
ストレージ装置
130
I/F
132
制御装置
134
メモリ
136
制御プログラム
138
RC プログラム
140
制御情報
142
144

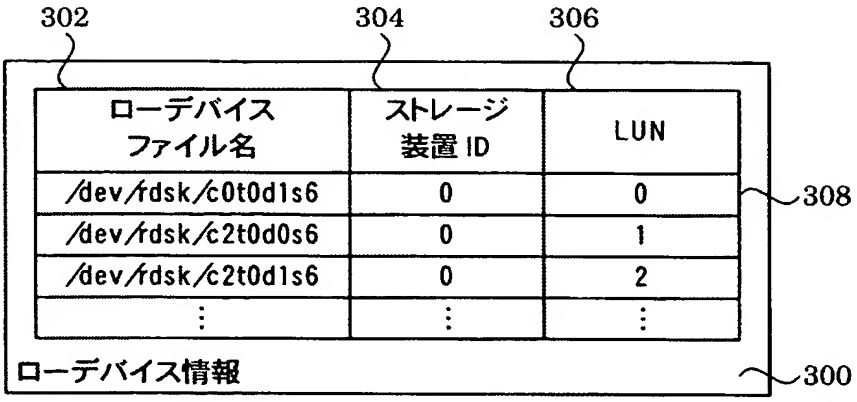
【図２】

図2


(a)
制御情報
142
LU 情報
200
グループ情報
202
(b)
210
212
214
LUN
ブロック数
サーバ ID
0
20971520
0
1
419430400
0
2
209715200
0
3
209715200
0
…
…
…
216
LU 情報
200
(c)
220
222
224
グループ ID
データ
優先度
LUN
0
0
0
1
1
1.4.…
2
5
2.5.…
3
3
3.6.…
…
…
…
226
グループ情報
202

【図３】

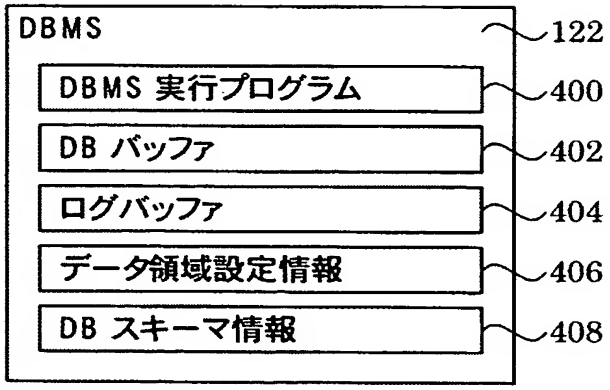
図3


302
304
306
ローデバイス
ファイル名
ストレージ
装置 ID
LUN
/dev/rdsk/c0t0d1s6
0
0
/dev/rdsk/c2t0d0s6
0
1
/dev/rdsk/c2t0d1s6
0
2
308
ローデバイス情報
300

【図４】

図4

(a)



(b)

410 データ領域 ID	412 データ領域名	414 ローデバイスファイル名	416 データ種別	418 領域サイズ	420 データ優先度
0	LOG	/dev/rdsk/c0t0d1s6	LOG	4GB	0
1	DB1	/dev/rdsk/c2t0d0s6	DATA	200GB	1
2	DB2	/dev/rdsk/c2t0d1s6	DATA	100GB	1
⋮	⋮	⋮	⋮	⋮	⋮

422

データ領域設定情報　406

(c)

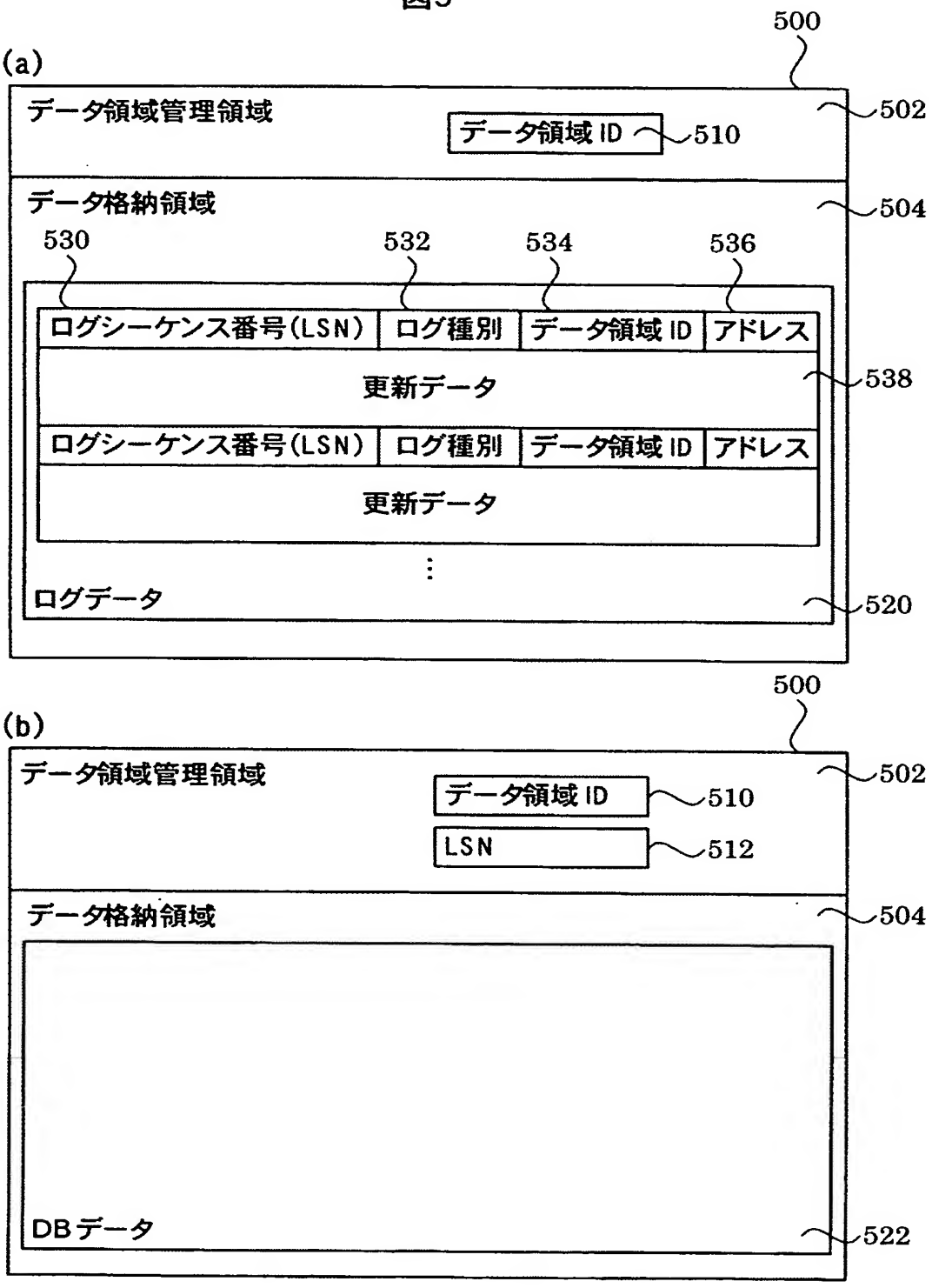
430 スキーマ ID	432 スキーマ名	434 スキーマ種別	436 データ領域 ID	438 スキーマサイズ
0	LOG1	LOG	0	2GB
1	TBL1	TABLE	1	100GB
2	IDX1	INDEX	2	50GB
⋮	⋮	⋮	⋮	⋮

440

DB スキーマ情報　408

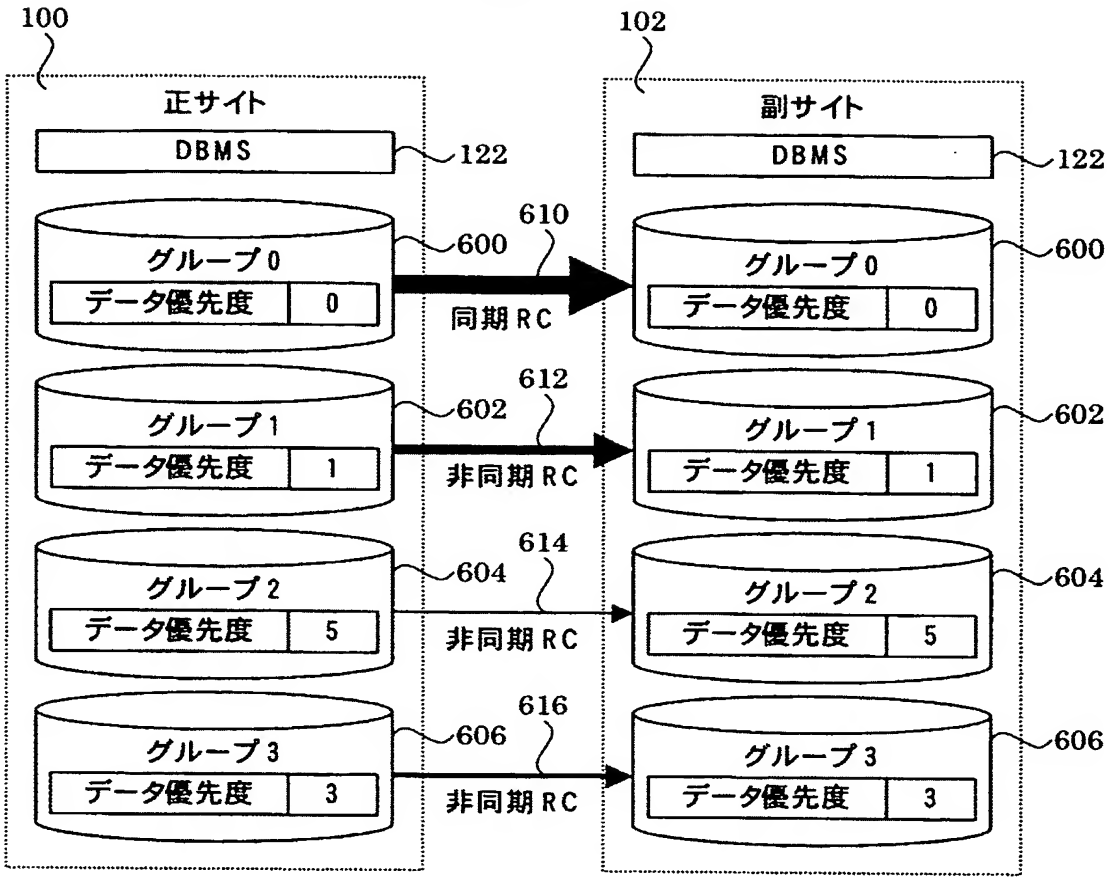
【図５】

図5


(a)
500
データ領域管理領域
データ領域ID
510
502
データ格納領域
504
530
532
534
536
ログシーケンス番号(LSN)
ログ種別
データ領域ID
アドレス
更新データ
538
ログシーケンス番号(LSN)
ログ種別
データ領域ID
アドレス
更新データ
ログデータ
520
(b)
500
データ領域管理領域
データ領域ID
510
LSN
512
502
データ格納領域
504
DBデータ
522

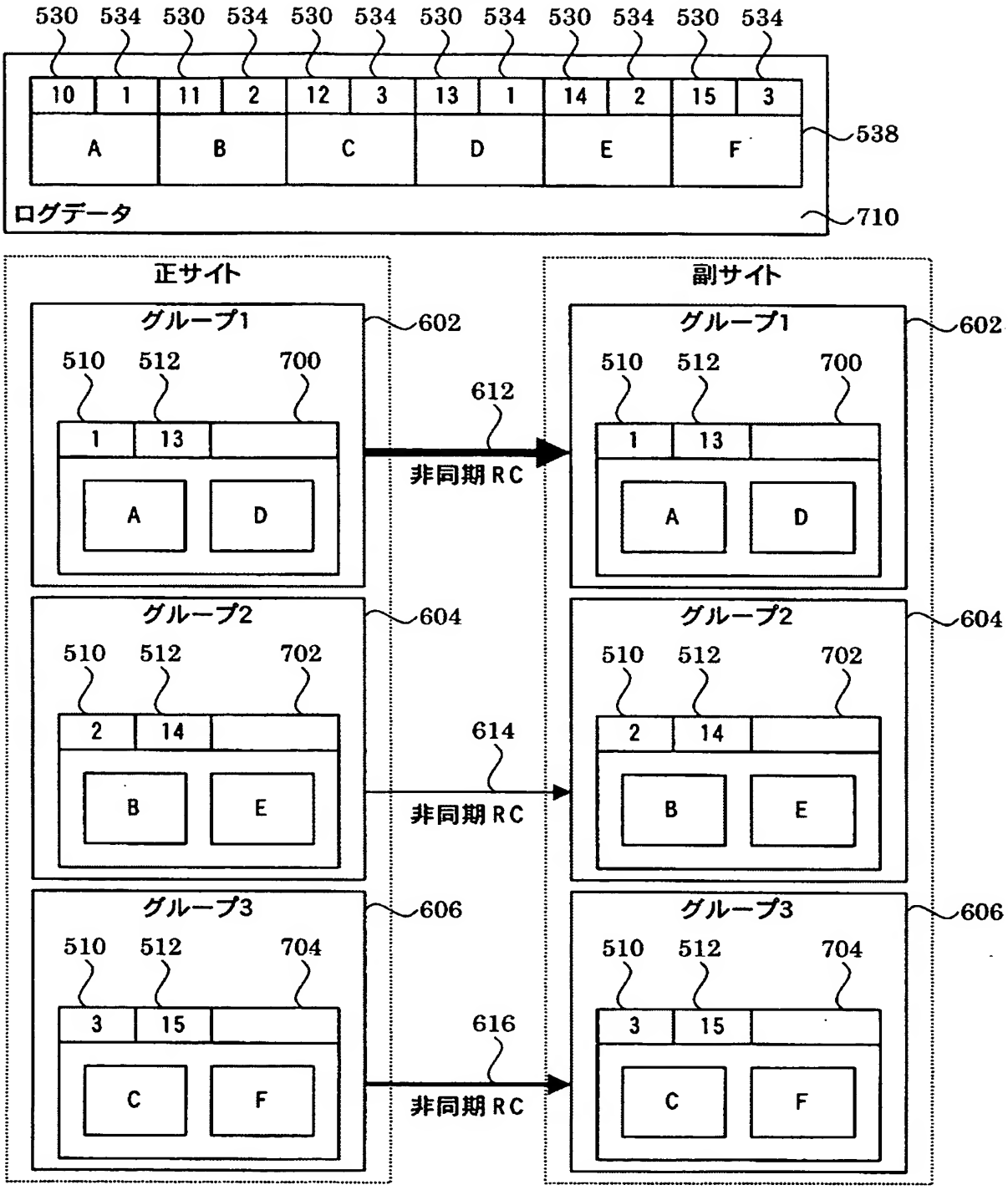
【図６】

図6


100
102
正サイト
副サイト
DBMS
122
DBMS
122
610
600
グループ0
データ優先度
0
同期RC
グループ0
データ優先度
0
600
612
602
グループ1
データ優先度
1
非同期RC
グループ1
データ優先度
1
602
614
604
グループ2
データ優先度
5
非同期RC
グループ2
データ優先度
5
604
616
606
グループ3
データ優先度
3
非同期RC
グループ3
データ優先度
3
606

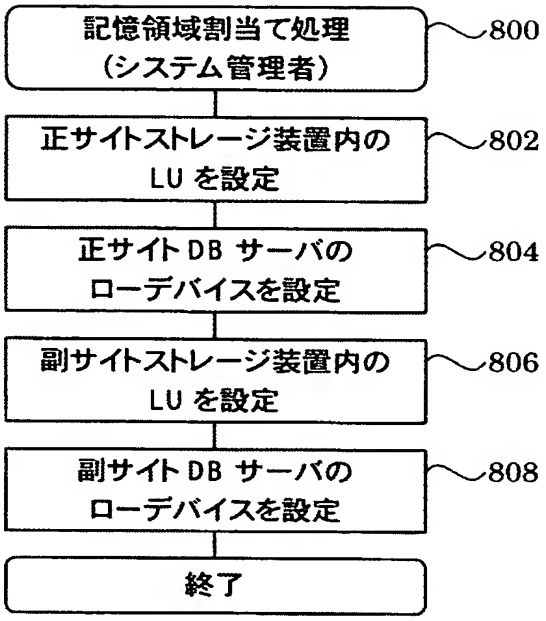
【図７】

図7


530 534 530 534 530 534 530 534 530 534 530 534
10 1 11 2 12 3 13 1 14 2 15 3
A B C D E F
538
ログデータ
710
正サイト
副サイト
グループ1
602
510 512 700
1 13
A D
612
非同期RC
グループ2
604
510 512 702
2 14
B E
614
非同期RC
グループ3
606
510 512 704
3 15
C F
616
非同期RC

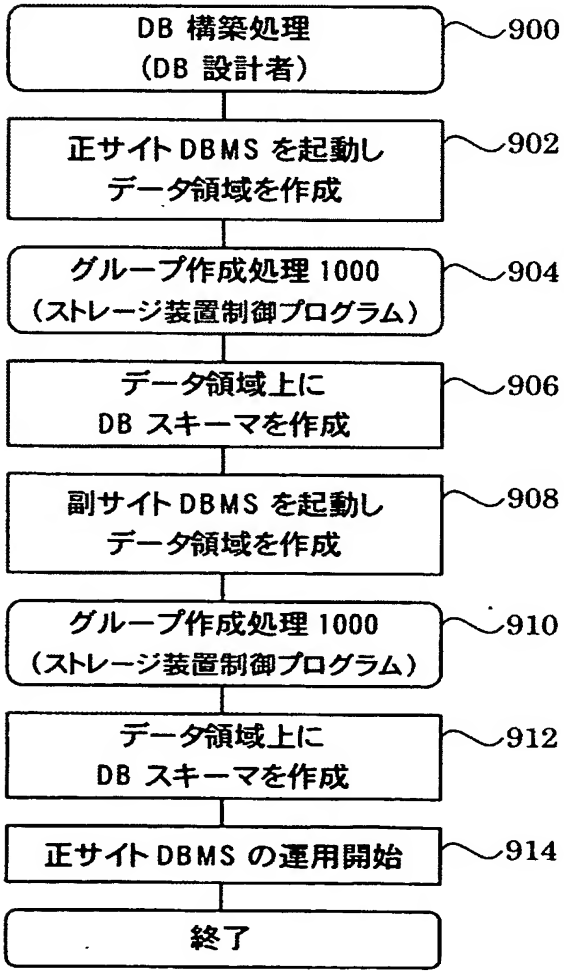
【図８】

図8


記憶領域割当て処理
（システム管理者）
800
正サイトストレージ装置内の
LU を設定
802
正サイト DB サーバの
ローデバイスを設定
804
副サイトストレージ装置内の
LU を設定
806
副サイト DB サーバの
ローデバイスを設定
808
終了

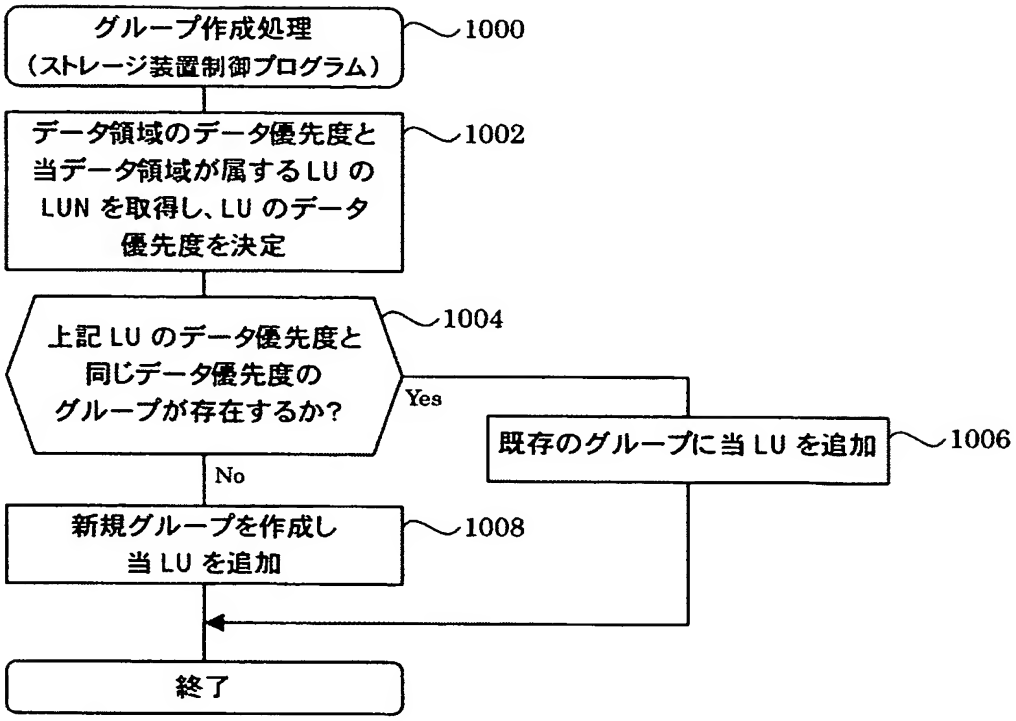
【図９】

図9


DB 構築処理
（DB 設計者）
900
正サイト DBMS を起動し
データ領域を作成
902
グループ作成処理 1000
（ストレージ装置制御プログラム）
904
データ領域上に
DB スキーマを作成
906
副サイト DBMS を起動し
データ領域を作成
908
グループ作成処理 1000
（ストレージ装置制御プログラム）
910
データ領域上に
DB スキーマを作成
912
正サイト DBMS の運用開始
914
終了

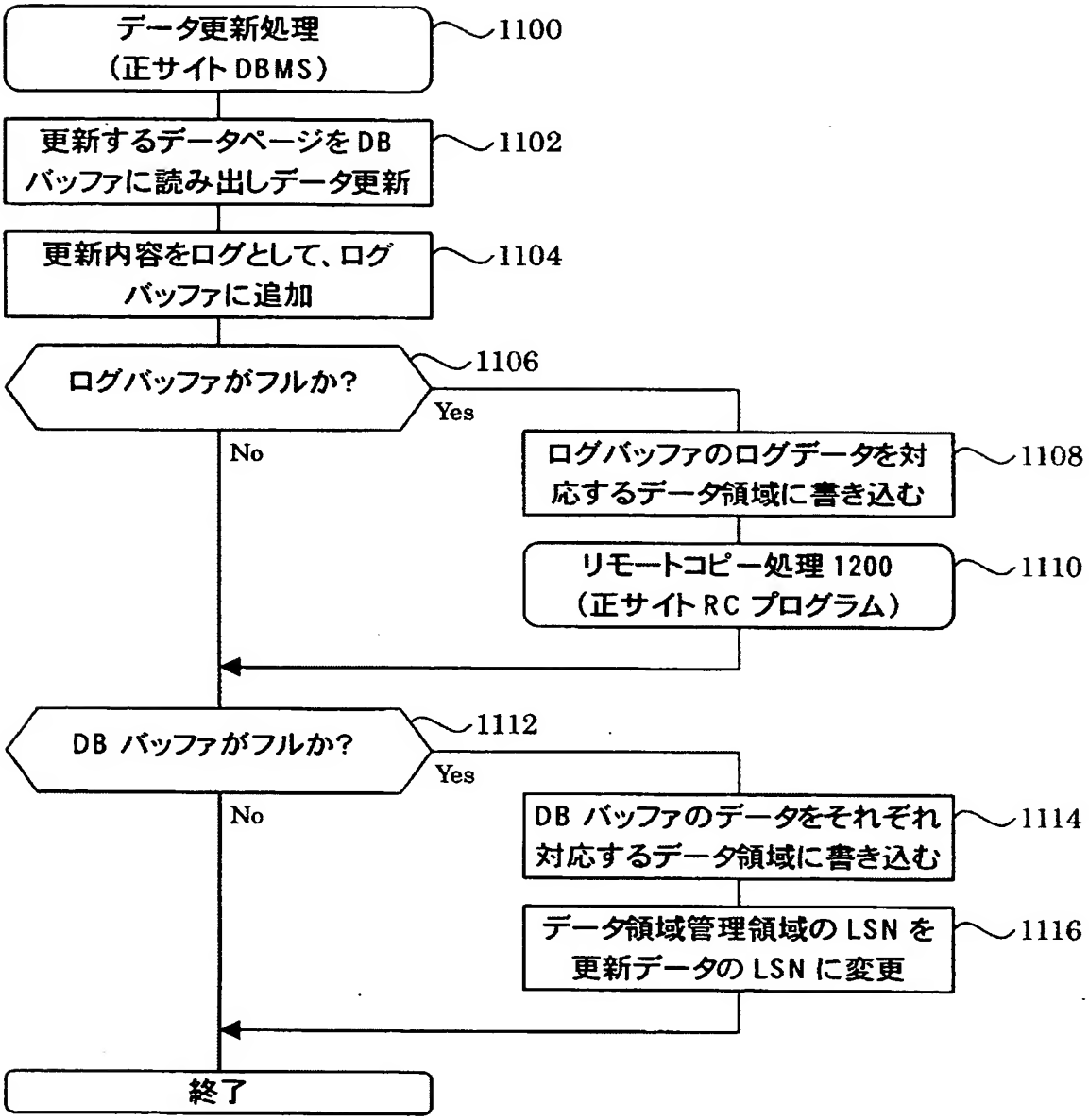
【図１０】

図10


グループ作成処理
（ストレージ装置制御プログラム）
1000
データ領域のデータ優先度と
当データ領域が属するLUの
LUNを取得し、LUのデータ
優先度を決定
1002
上記LUのデータ優先度と
同じデータ優先度の
グループが存在するか?
1004
Yes
既存のグループに当LUを追加
1006
No
新規グループを作成し
当LUを追加
1008
終了

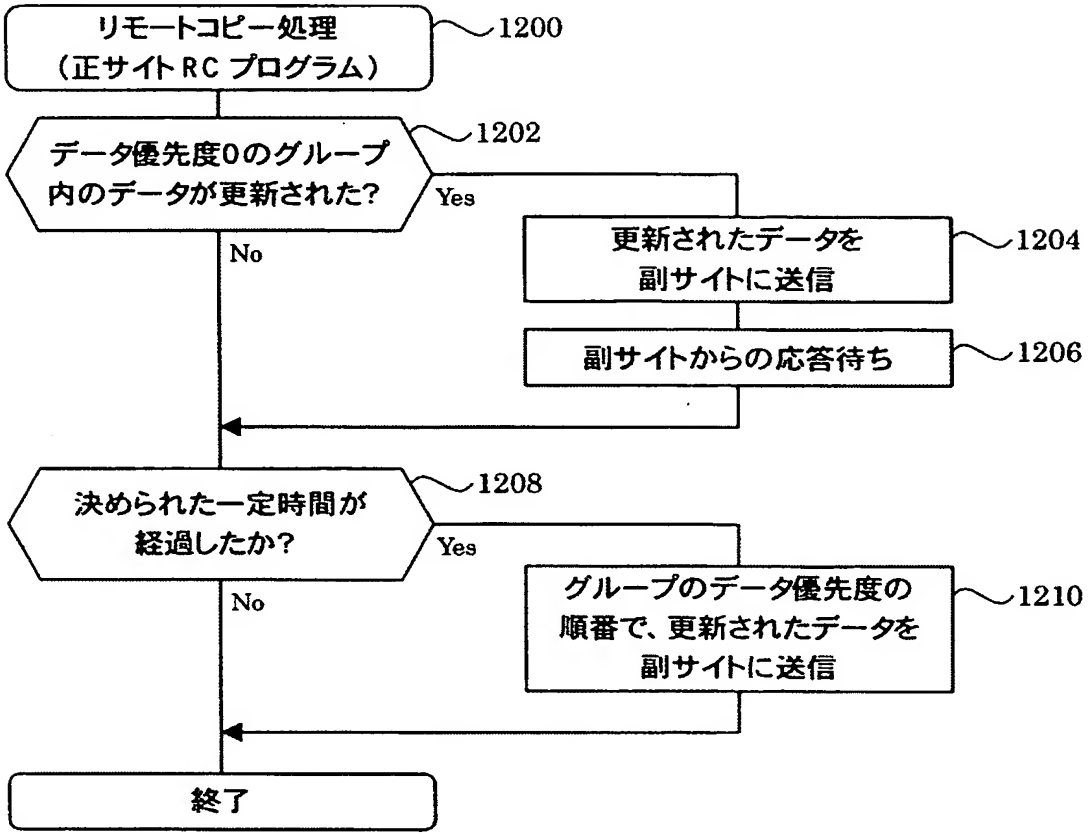
【図１１】

図11


データ更新処理
（正サイトDBMS）
1100
更新するデータページをDB
バッファに読み出しデータ更新
1102
更新内容をログとして、ログ
バッファに追加
1104
ログバッファがフルか?
1106
Yes
No
ログバッファのログデータを対
応するデータ領域に書き込む
1108
リモートコピー処理1200
（正サイトRCプログラム）
1110
DBバッファがフルか?
1112
Yes
No
DBバッファのデータをそれぞれ
対応するデータ領域に書き込む
1114
データ領域管理領域のLSNを
更新データのLSNに変更
1116
終了

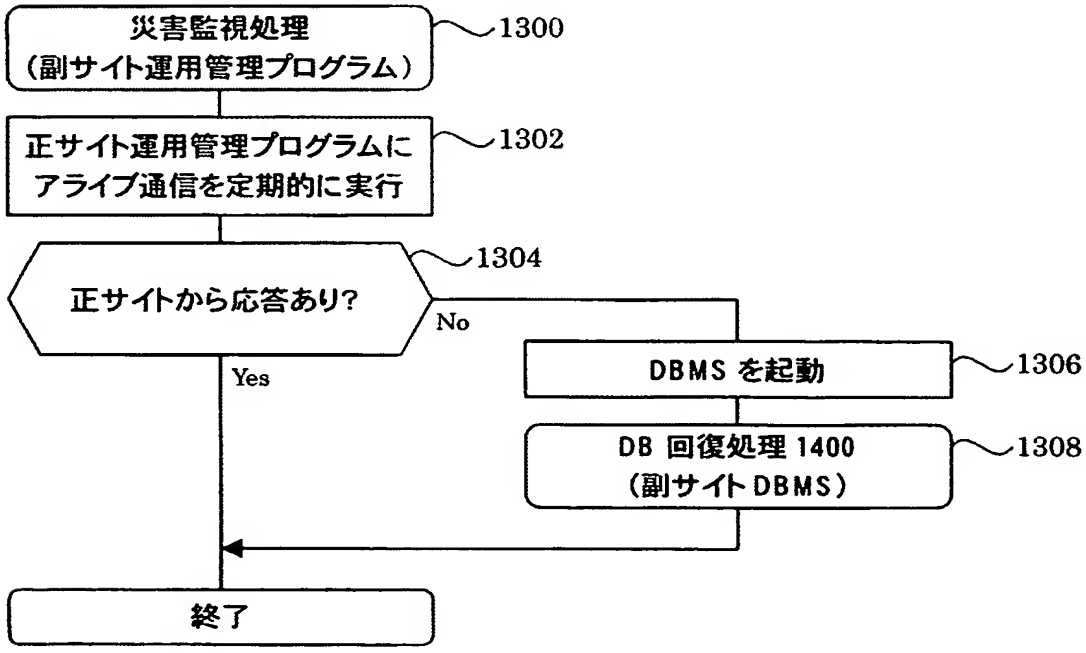
【図１２】

図12



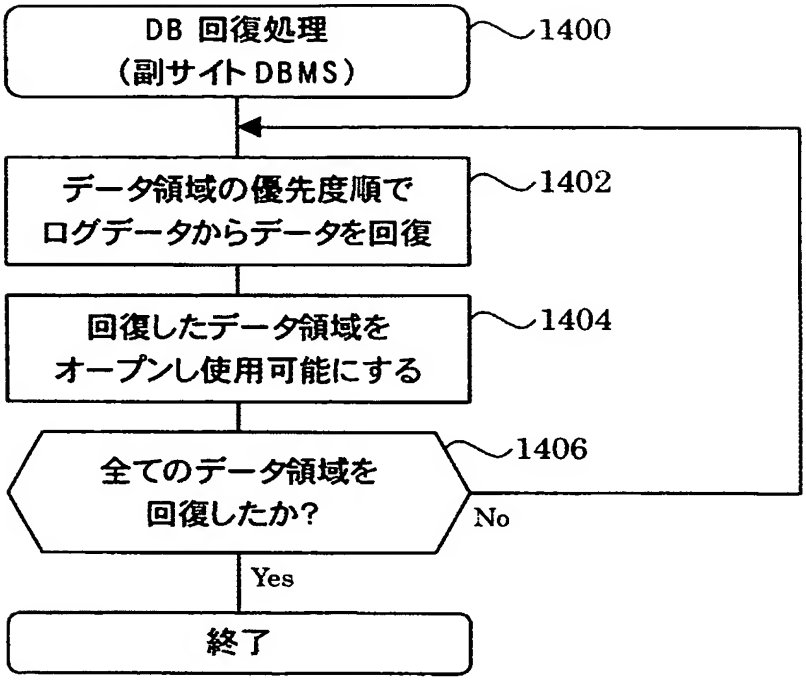
【図１３】

図13



【図１４】

図14


DB 回復処理
（副サイトDBMS）
1400
データ領域の優先度順で
ログデータからデータを回復
1402
回復したデータ領域を
オープンし使用可能にする
1404
全てのデータ領域を
回復したか?
1406
No
Yes
終了

【書類名】要約書
【要約】
【課題】
ＤＲシステムとＤＢを構築する際、ＤＢデータ毎に復旧時間の要求が異なる場合も、従来技術ではそれらを考慮することはなく、最も条件の高いＤＢデータに対する要求がＤＲシステム構築の要件となり、システムがオーバースペックになる可能性が高いという課題がある。
【解決手段】
本発明では、ある管理単位毎にＤＢデータの優先度を設定する手段、そのデータ優先度順に正サイトから副サイトにデータをコピーする手段、正サイトに災害が発生し副サイトでＤＢの回復を行う際、データ優先度順にデータの回復を行い、順次使用可能な状態にする手段を設ける。
【選択図】　図６

認定・付加情報

特許出願の番号	特願２００４－０４７１７６
受付番号	５０４００２９０３７０
書類名	特許願
担当官	末武　実　　１９１２
作成日	平成１６年　２月２５日

＜認定情報・付加情報＞

【提出日】　平成16年 2月24日

特願２００４－０４７１７６

出　願　人　履　歴　情　報

識別番号　　［０００００５１０８］

１．変更年月日　　１９９０年　８月３１日
［変更理由］　　新規登録
住　所　　東京都千代田区神田駿河台４丁目６番地
氏　名　　株式会社日立製作所